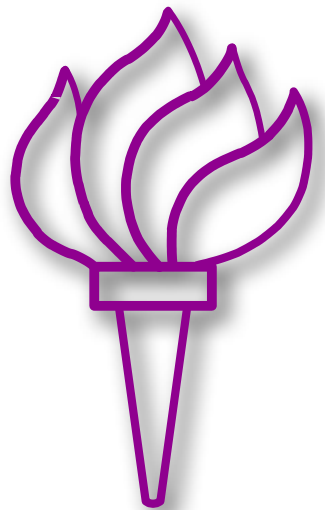
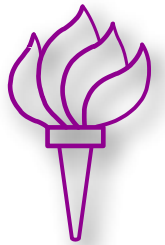
# PROOF installation/usage



Attila Krasznahorkay
for the Tier3 PROOF WG

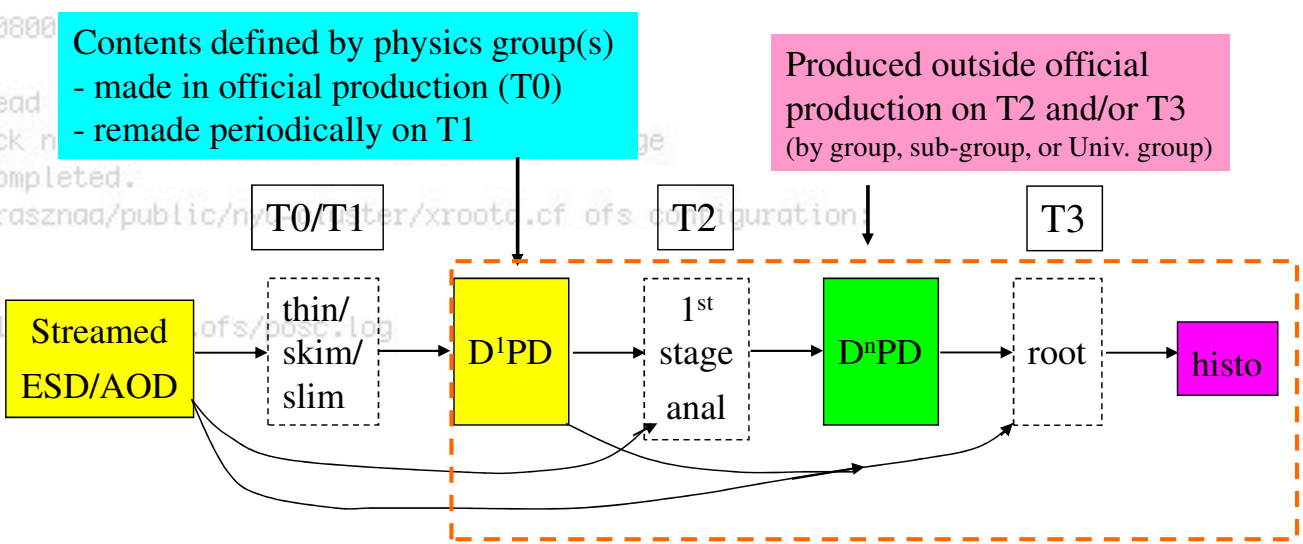
# Overview



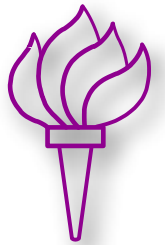
- PROOF recap
- The work done in the WG
- WG Recommendations
  - PROOF installation/configuration
  - Using PROOF efficiently
  - Usage of PQ2
  - Usage of SFrame

# Analysis model with D$^n$PDs

- Users are encouraged to use D3PDs (simple ROOT ntuples) for analysis
  - Small dataset sizes
  - Quick processing of events
- D3PDs are created either on Tier2-s or Tier3-s



# Processing D3PDs

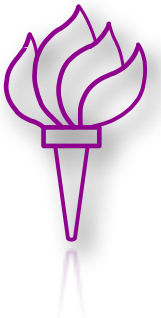


- Current D3PD sizes: up to 20 kb/event
  - People will need to process multiple TBs of data with quick turnaround soon
- Single-core analyses: up to few kHz event processing rate
  - Processing "just" 20M data events takes a few hours
  - We already have more than this in some analyses
- Have to run the ROOT jobs in parallel

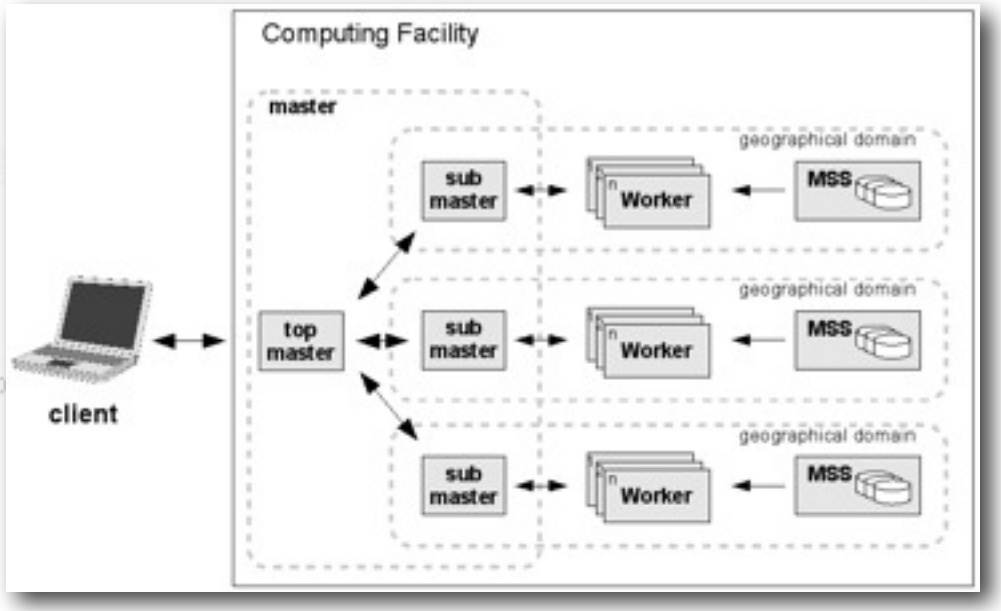
# General PROOF concepts

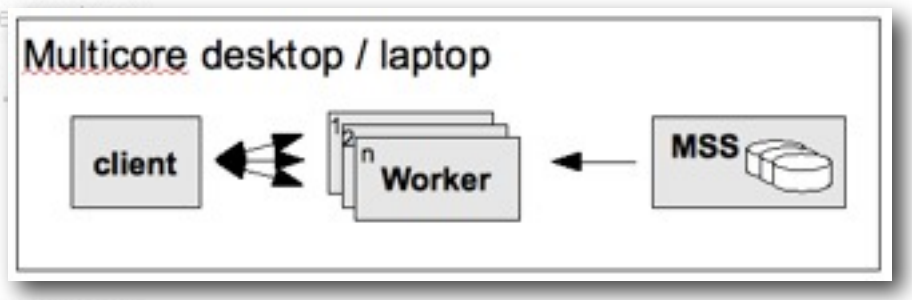
```
100602 18:15:01 001 Scalla is starting. . .
Copr.  2007 Stanford University, xrd version v20100205-0000-root-4
++++++ xrootd anon@pcnyu02.cern.ch initialization started.
Config using configuration file /afs/cern.ch/user/k/krasznaa/public/nyu-cluster/xrootd.cf
=====> xrd.port 1094
=====> xrd.protocol xproofd:1093 /usr/local/root/lib/libXrdProofd.so
Config maximum number of connections restricted to 1024
100602 18:15:01 001 xrd.protocol xproofd:1093 /usr/local/root/lib/libXrdProofd.so
Copr.  2007 Stanford University, xrootd version 2.9.1 build v20100205-0000-root-4
++++++ xrootd protocol initialization started.
=====> xrootd.fslib /usr/local/root/lib/libXrdOfs.so
=====> all.export /home/proof r/w
Config exporting /home/proof
Config warning: 'xrootd.seclib' not specified; strong authentication disabled!
Copr.  2008 Stanford University, Ofs Version v20100205-0000-root-4
++++++ File system initialization started.
Copr.  2008, Stanford University, oss Version v20100205-0000-root-4
++++++ Storage system initialization started.
=====> all.export /home/proof r/w
Config effective /afs/cern.ch/user/k/krasznaa/public/nyu-cluster/xrootd.cf oss configuration:
       oss.alloc       0 0 0
       oss.cachescan   600
       oss.compdetect  *
       oss.fdlimit     512 1024
       oss.maxdbsize   0
       oss.trace       0
       oss.xfr         1 9437184 30 10800
       oss.memfile off  max 1931927552
       oss.defaults  r/w  nocheck nodread nomig norcreate nopurge nostage
       oss.path /home/proof r/w  nocheck nodread nomig norcreate nopurge nostage
------ Storage system initialization completed.
Config effective /afs/cern.ch/user/k/krasznaa/public/nyu-cluster/xrootd.cf ofs configuration:
       ofs.role server
       ofs.maxdelay   60
       ofs.persist    manual hold 600 logdir /tmp/.ofs/posc.log
       ofs.trace      0
------ File system server initialization completed.
Config warning: 'xrootd.prepare logdir' not specified; prepare tracking disabled.
------ xrootd protocol initialization completed.
100602 18:15:01 001 xrd.protocol xproofd:1093 /usr/local/root/lib/libXrdProofd.so
100602 18:15:01 001 xpd.workdir /home/proof/
100602 18:15:01 001 xpd.maxoldlogs 5
100602 18:15:01 001 xpd.role worker
```

# PROOF - what is it?

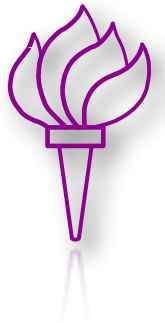


- Lot of information on the ROOT webpage: http://root.cern.ch/drupal/content/proof
- Also, multiple presentations already: http://indico.cern.ch/getFile.py/access?contribId=19&resId=3&materialId=slides&confId=71202



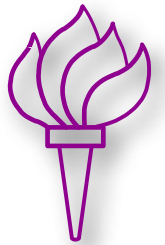



# PROOF - features



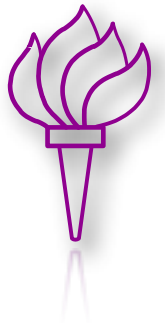
- Main advantages:
  - Only a recent ROOT installation needed
  - Can connect workers of different architecture
  - Job splitting is optimised (slower workers process less events)
  - Scalable way beyond the Tier3 needs
  - Provides easy to use interfaces, hides the complexity of the system
  - Can be used interactively
  - Output merging is handled by ROOT
  - PROOF-Lite provides a zero-configuration setup for running jobs on all cores of a single machine

# PROOF - requirements



- Needs a storage system for the input of the jobs
  - Can be any system in principle (as long as `TFile::Open(...)` supports it, it's fine)
    - XRootD - preferred for many reasons
    - dCache
    - Lustre
    - gpfs
    - Castor
    - ...
  - The performance of the storage system pretty much defines the performance of the PROOF cluster

# The working group

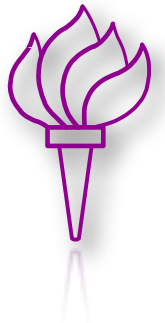


- Main TWiki page:
  https://twiki.cern.ch/twiki/bin/view/Atlas/AtlasProofWG
- Tasks:
  - Survey and evaluate current PROOF tools
  - Give instructions for Tier3 PROOF farm installations
  - Provide dataset management tools
  - Formulate Tier3 analysis best practices

# Setting up a PROOF cluster

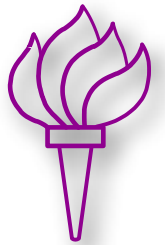
```
100602 18:15:01 001 Scalla is starting. . .
Copr.  2007 Stanford University, xrd version v20100205-0000-root-4
++++++ xrootd anon@pcnyu02.cern.ch initialization started.
Config using configuration file /afs/cern.ch/user/k/krasznaa/public/nyu-cluster/xrootd.cf
=====> xrd.port 1094
=====> xrd.protocol xproofd:1093 /usr/local/root/lib/libXrdProofd.so
Config maximum number of connections restricted to 1024
100602 18:15:01 001 xrd.protocol xproofd:1093 /usr/local/root/lib/libXrdProofd.so
Copr.  2007 Stanford University, xrootd version 2.9.1 build v20100205-0000-root-4
++++++ xrootd protocol initialization started.
=====> xrootd.fslib /usr/local/root/lib/libXrdOfs.so
=====> all.export /home/proof r/w
Config exporting /home/proof
Config warning: 'xrootd.seclib' not specified; strong authentication disabled!
Copr.  2008 Stanford University, Ofs Version v20100205-0000-root-4
++++++ File system initialization started.
Copr.  2008, Stanford University, oss Version v20100205-0000-root-4
++++++ Storage system initialization started.
=====> all.export /home/proof r/w
Config effective /afs/cern.ch/user/k/krasznaa/public/nyu-cluster/xrootd.cf oss configuration:
       oss.alloc       0 0 0
       oss.cachescan   600
       oss.compdetect  *
       oss.fdlimit     512 1024
       oss.maxdbsize   0
       oss.trace       0
       oss.xfr         1 9437184 30 10800
       oss.memfile off  max 1931927552
       oss.defaults  r/w  nocheck nodread nomig norcreate nopurge nostage
       oss.path /home/proof r/w  nocheck nodread nomig norcreate nopurge nostage
------ Storage system initialization completed.
Config effective /afs/cern.ch/user/k/krasznaa/public/nyu-cluster/xrootd.cf ofs configuration:
       ofs.role server
       ofs.maxdelay   60
       ofs.persist    manual hold 600 logdir /tmp/.ofs/posc.log
       ofs.trace      0
------ File system server initialization completed.
Config warning: 'xrootd.prepare logdir' not specified; prepare tracking disabled.
------ xrootd protocol initialization completed.
100602 18:15:01 001 xrd.protocol xproofd:1093 /usr/local/root/lib/libXrdProofd.so
100602 18:15:01 001 xpd.workdir /home/proof/
100602 18:15:01 001 xpd.maxoldlogs 5
100602 18:15:01 001 xpd.role worker
```

# Installation



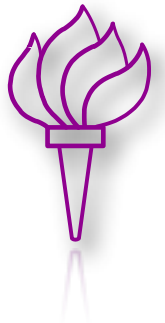
- Special tag of ROOT created for this: http://root.cern.ch/drupal/content/root-version-v5-26-00-proof
- Includes some improvements over ROOT version 5.26, plus all the newest PQ2 tools
- Installation is summarised on: https://twiki.cern.ch/twiki/bin/view/Atlas/HowToInstallPROOFWithXrootdSystem
- Storage system installation/setup is not covered

# Configuration



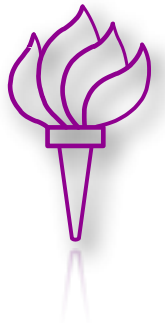
- Configuration file uses the same syntax as XRootD
  - Most common configuration of PROOF is to run the PROOF executable by xrootd
- The recommended installation uses the xrootd daemon packaged with the recommended version of ROOT
  - Example configuration file provided on the TWiki
  - Needs some expert knowledge to fine tune at the moment

# PROOF and XRootD



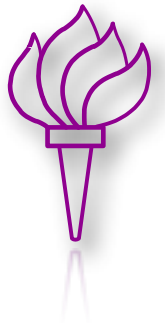
- PROOF needs some XRootD shares to work properly
  - When writing large outputs, each worker node has to export its workarea using xrootd for the PROOF master node
  - There has to be a scratch area that the master node can write, and the client node can read (for the merged output files)
- Usually PROOF and XRootD are set up using a single configuration file -> Poses a possible overhead if we don't do it at Tier3-s

# PROOF on a batch (1)



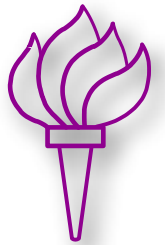
- In most cases the PROOF cluster uses the same worker nodes as the batch cluster, running the daemons in parallel
  - For small clusters/groups this is usually not a problem -> resources are shared after discussion among the users
  - Larger sites should do something more sophisticated
- The batch cluster can be made aware of the PROOF daemon, holding back the batch jobs while PROOF jobs complete

# PROOF on a batch (2)



- PROOF on Demand (PoD, http://pod.gsi.de):
  - Submits jobs to the batch cluster, running the PROOF master and worker processes as user programs
  - Can use the batch system to balance resources between users
  - Developed at GSI, used there with big success
  - No backend for Condor yet, but could possibly convince the developer of providing one
  - No robust support for the project at the moment (personal impression)

# Monitoring

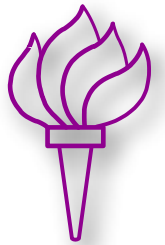


- Can use Ganglia, just like for XRootD monitoring
  - Started the documentation on: https://twiki.cern.ch/twiki/bin/view/Atlas/MonitoringAPROOFCluster
- The monitoring of jobs can be done using MonAlisa (http://monalisa.caltech.edu)
  - The recommended ROOT binary comes with the MonAlisa libraries linked in
  - Developed for the ALICE collaboration, but general enough to be used by ATLAS
  - No good instructions for the setup yet

```
100602 18:15:01 001 Scalla is starting. . .
Copr.  2007 Stanford University, xrd version v20100205-0000-root-4
++++++ xrootd anon@pcnyu02.cern.ch initialization started.
Config using configuration file /afs/cern.ch/user/k/krasznaa/public/nyu-cluster/xrootd.cf
=====> xrd.port 1094
=====> xrd.protocol xproofd:1093 /usr/local/root/lib/libXrdProofd.so
Config maximum number of connections restricted to 1024
100602 18:15:01 001 xrd.protocol xproofd:1093 /usr/local/root/lib/libXrdProofd.so
Copr.  2007 Stanford University, xrootd version 2.9.1 build v20100205-0000-root-4
++++++ xrootd protocol initialization started.
=====> xrootd.fslib /usr/local/root/lib/libXrdOfs.so
=====> all.export /home/proof r/w
Config exporting /home/proof
Config warning: 'xrootd.seclib' not specified; strong authentication disabled!
Copr.  2008 Stanford University, Ofs Version v20100205-0000-root-4
++++++ File system initialization started.
Copr.  2008, Stanford University, oss Version v20100205-0000-root-4
++++++ Storage system initialization started.
=====> all.export /home/proof r/w
Config effective /afs/cern.ch/user/k/krasznaa/public/nyu-cluster/xrootd.cf oss configuration:
       oss.alloc       0 0
       oss.cachescan   600
       oss.compdetect  *
       oss.fdlimit     512 1024
       oss.maxdbsize   0
       oss.trace       0
       oss.xfr         1 9437184 30 10800
       oss.memfile off  max 1931927552
       oss.defaults  r/w  nocheck nodread nomig norcreate nopurge nostage
       oss.path /home/proof r/w  nocheck nodread nomig norcreate nopurge nostage
------ Storage system initialization completed.
Config effective /afs/cern.ch/user/k/krasznaa/public/nyu-cluster/xrootd.cf ofs configuration:
       ofs.role server
       ofs.maxdelay   60
       ofs.persist    manual hold 600 logdir /tmp/.ofs/posc.log
       ofs.trace      0
------ File system server initialization completed.
Config warning: 'xrootd.prepare logdir' not specified; prepare tracking disabled.
------ xrootd protocol initialization completed.
100602 18:15:01 001 xrd.protocol xproofd:1093 /usr/local/root/lib/libXrdProofd.so
100602 18:15:01 001 xpd.workdir /home/proof/
100602 18:15:01 001 xpd.maxoldlogs 5
100602 18:15:01 001 xpd.role worker
```

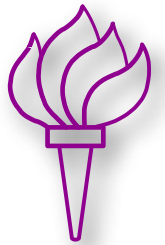
# Handling datasets

# Dataset management



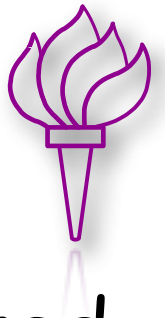
- A set of scripts (PQ2) are provided to manage datasets on PROOF farms
  - Very similar to DQ2 (hence the name...)
  - Users don't have to know the location of each file, they can run the PROOF jobs on the named datasets
- Basic documentation is here:
  http://root.cern.ch/drupal/content/pq2-tools

# Dataset management



- Description of registering a DQ2 dataset on PQ2 is available here:
  https://twiki.cern.ch/twiki/bin/view/Atlas/HowToUsePQ2ToManageTheLocalDatasets
  - Download the dataset into a temporary directory with dq2-get
  - Copy the files onto the XRootD redirector with xrdcp, while creating a local file list
  - Register the dataset using pq2-put with the local file list
  - Management only done by site administrators

# Dataset usage



- Users can get information from the registered datasets with the PQ2 tools

```
> pq2-ls
Dataset repository: /home/proof/krasznaa/datasets
Dataset URI                                  | # Files | Default tree | # Events |    Disk    | Staged
/default/krasznaa/SFrameTestDataSet          |       1 | /CollectionT>| 1.25e+04 |   148 MB   |  100 %
/default/krasznaa/SFrameTestDataSet2         |       1 | /CollectionT>| 1.25e+04 |   148 MB   |  100 %
/default/krasznaa/data10_7TeV.00153030.physics_MinBias.merge.NTUP_EGAM.f247_p129|     726 | /CollectionT>|
4.006e+06|    13 GB |  100 %

> pq2-ls-files /default/krasznaa/data10_7TeV.00153030.physics_MinBias.merge.NTUP_EGAM.f247_p129
pq2-ls-files: dataset '/default/krasznaa/data10_7TeV.00153030.physics_MinBias.merge.NTUP_EGAM.f247_p129'
has 726 files
pq2-ls-files: #   File                                                                    Size  #Objs
Obj|Type|Entries, ...
pq2-ls-files: 1   root://krasznaa@//pool0/data10_7TeV/NTUP_EGAM/data10_7TeV.
00153030.physics_MinBias.merge.NTUP_EGAM.f247_p129_tid126434_00/NTUP_EGAM.126434._000001.root.1 35 MB  2
CollectionTree|TTree|10923,egamma|TTree|10923
pq2-ls-files: 2   root://krasznaa@//pool0/data10_7TeV/NTUP_EGAM/data10_7TeV.
00153030.physics_MinBias.merge.NTUP_EGAM.f247_p129_tid126434_00/NTUP_EGAM.126434._000002.root.1 34 MB  2
CollectionTree|TTree|10647,egamma|TTree|10647
pq2-ls-files: 3   root://krasznaa@//pool0/data10_7TeV/NTUP_EGAM/data10_7TeV.
00153030.physics_MinBias.merge.NTUP_EGAM.f247_p129_tid126434_00/NTUP_EGAM.126434._000003.root.1  8 MB  2
CollectionTree|TTree|2611,egamma|TTree|2611
...
```

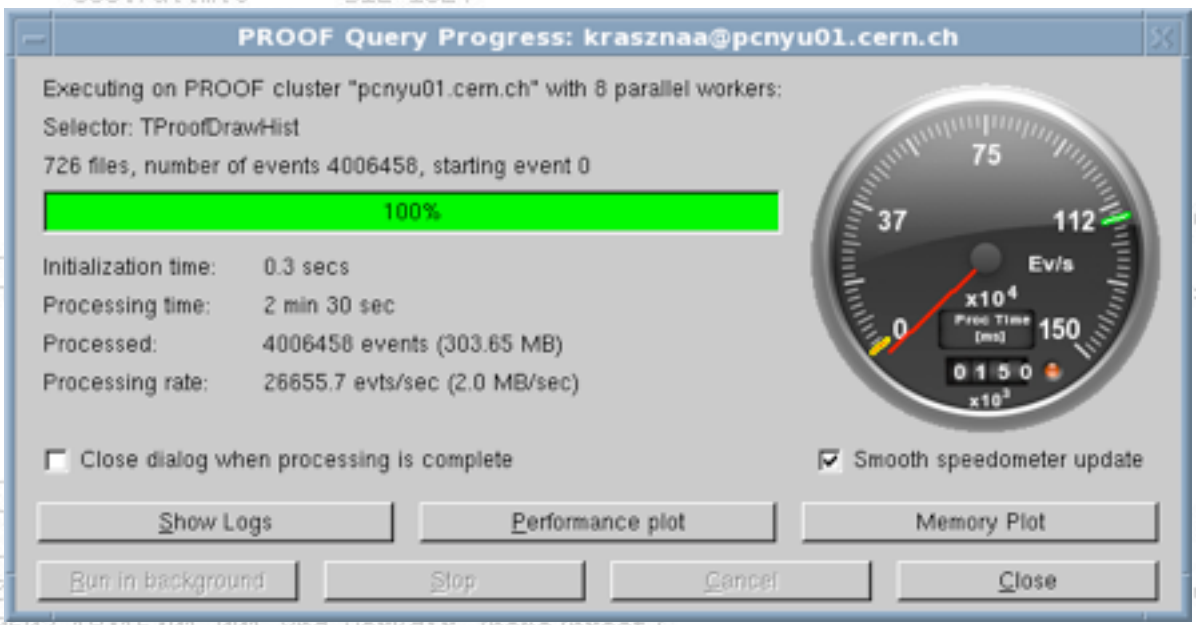
```
100602 18:15:01 001 Scalla is starting. . .
Copr.  2007 Stanford University, xrd version v20100205-0000-root-4
++++++ xrootd anon@pcnyu02.cern.ch initialization started.
Config using configuration file /afs/cern.ch/user/k/krasznaa/public/nyu-cluster/xrootd.cf
=====> xrd.port 1094
=====> xrd.protocol xproofd:1093 /usr/local/root/lib/libXrdProofd.so
Config maximum number of connections restricted to 1024
100602 18:15:01 001 xrd.protocol xproofd:1093 /usr/local/root/lib/libXrdProofd.so
Copr.  2007 Stanford University, xrootd version 2.9.1 build v20100205-0000-root-4
++++++ xrootd protocol initialization started.
=====> xrootd.fslib /usr/local/root/lib/libXrdOfs.so
=====> all.export /home/proof r/w
Config exporting /home/proof
Config warning: 'xrootd.seclib' not specified; strong authentication disabled!
Copr.  2008 Stanford University, Ofs Version v20100205-0000-root-4
++++++ File system initialization started.
Copr.  2008, Stanford University, oss Version v20100205-0000-root-4
++++++ Storage system initialization started.
=====> all.export /home/proof r/w
Config effective /afs/cern.ch/user/k/krasznaa/public/nyu-cluster/xrootd.cf oss configuration:
       oss.alloc       0 0 0
       oss.cachescan   600
       oss.compdetect  *
       oss.fdlimit     512 1024
       oss.maxdbsize   0
       oss.trace       0
       oss.xfr         1 9437184 30 10800
       oss.memfile off  max 1931927552
       oss.defaults  r/w  nocheck nodread nomig norcreate nopurge nostage
       oss.path /home/proof r/w  nocheck nodread nomig norcreate nopurge nostage
------ Storage system initialization completed.
Config effective /afs/cern.ch/user/k/krasznaa/public/nyu-cluster/xrootd.cf ofs configuration:
       ofs.role server
       ofs.maxdelay   60
       ofs.persist    manual hold 600 logdir /tmp/.ofs/posc.log
       ofs.trace      0
------ File system server initialization completed.
Config warning: 'xrootd.prepare logdir' not specified; prepare tracking disabled.
------ xrootd protocol initialization completed.
100602 18:15:01 001 xrd.protocol xproofd:1093 /usr/local/root/lib/libXrdProofd.so
100602 18:15:01 001 xpd.workdir /home/proof/
100602 18:15:01 001 xpd.maxoldlogs 5
100602 18:15:01 001 xpd.role worker
```

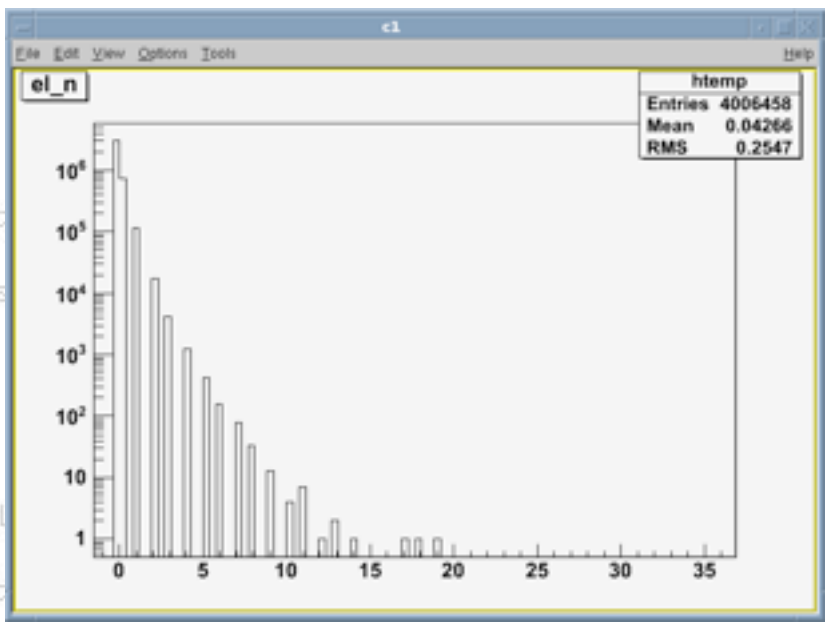
# Running jobs

# Using PROOF

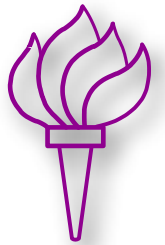
- Simplest use case: In interactive mode

```
root [0] p = TProof::Open( “username@master.domain.edu” );
Starting master: opening connection ...
Starting master: OK
Opening connections to workers: OK (XX workers)
Setting up worker servers: OK (XX workers)
PROOF set to parallel mode (XX workers)
root [1] p->DrawSelect( “/default/dataset#egamma”, “el_n” );
```



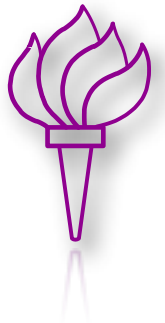



# Using PROOF



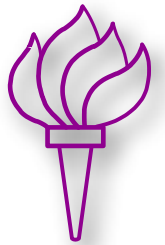
- The user can write his/her analysis code using the TSelector class
  - The base class provides the virtual functions that are called during the event loop
- Documentation is available here:
  http://root.cern.ch/drupal/content/developing-tselector
- Benchmark example created by the WG is here:
  https://twiki.cern.ch/twiki/bin/view/Atlas/BenchmarksWithDifferentConfigurations#Native_PROOF_example

# Using PROOF



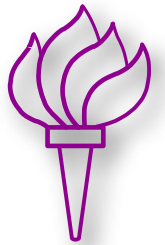
- Full-scale analyses can be written using SFrame
  - Main documentation: http://sframe.sourceforge.net, http://sourceforge.net/apps/mediawiki/sframe/
  - Previous presentation: http://indico.cern.ch/getFile.py/access?contribId=13&resId=0&materialId=slides&confId=71202
- Example benchmark code given by the WG: https://twiki.cern.ch/twiki/bin/view/Atlas/BenchmarksWithDifferentConfigurations#SFrame_example

# SFrame continued



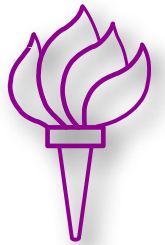
- Provides a framework for writing analysis package hierarchies
  - The framework takes care of packaging up the user code, distributing it to the worker nodes, and compiling it on each of them
- Gives a flexible configuration system for the jobs
  - Can run the jobs locally or using PROOF-Lite for debugging, then send the job to the PROOF cluster by just changing a configuration parameter

# SFrameARA



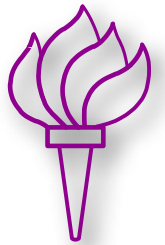
- SFrame can also analyse POOL files using ARA
  - Implemented as an extension to the ROOT-only SFrame code
- Can not use a proper PROOF cluster for processing POOL files (at the moment), but is able to use PROOF-Lite
  - ARA analysis jobs can run with speeds close to the ones produced by ntuple analyses
  - A number of people using it for serious analyses already

# Missing pieces



- The cluster configuration with the current instructions still needs some expert knowledge
  - Will have to agree on a model configuration for an average T3g
  - Have to come up with a method of helping the Tier3 administrators set up their systems (who will do it?)
- Recommended ROOT version distribution/ update not solved yet
  - Once the new features get into the main development branch, the next usual ROOT release will be fine as well

# Summary



- PROOF installation still needs expert knowledge
- Using a well configured PROOF cluster is relatively easy from the user perspective
  - Documentation is very good for ROOT
  - SFrame documented to a quite good degree, with multiple examples
- I/O is the main limiting factor - If the I/O can keep up, the speed increase is linear with the number of processor cores